Detection of meibomian glands and classification of meibography images

Yang Wei Koh
Turgay Celik
Hwee Kuan Lee
Andrea Petznick
Louis Tong
Detection of meibomian glands and classification of meibography images

Yang Wei Koh, Turgay Celik, Hwee Kuan Lee, Andrea Petznick, and Louis Tong

Abstract. Computational methods are presented that can automatically detect the length and width of meibomian glands imaged by infrared meibography without requiring any input from the user. The images are then automatically classified. The length of the glands are detected by first normalizing the pixel intensity, extracting stationary points, and then applying morphological operations. Gland widths are detected using scale invariant feature transform and analyzed using Shannon entropy. Features based on the gland lengths and widths are then used to train a linear classifier to accurately differentiate between healthy (specificity 96.1%) and unhealthy (sensitivity 97.9%) meibography images. The user-free computational method is fast, does not suffer from inter-observer variability, and can be useful in clinical studies where large number of images needs to be analyzed efficiently.

Keywords: dry-eye; meibography; diagnosis; computer vision; image processing; machine learning.

1 Introduction

Clinical research on dry-eye syndrome is currently receiving much attention. In particular, in the study of meibomian gland dysfunction, which is a major cause of dry-eye gland dysfunction, can be diagnosed by directly observing the morphology of the meibomian glands using meibography. Recently, a noncontact, patient-friendly infrared meibography method was developed that is fast and greatly reduced discomfort to the patient. This advance allows meibography to be used routinely in clinical research.

When analyzing meibography images, it is important to quantify the area of meibomian gland loss. Grades are then assigned to the image based on the area of gland loss, which in turn serve as an indicator of meibomian gland dysfunction. The loss of meibomian glands in the upper and lower eyelids have been studied by Arita et al., Pult et al., and Srinivasan et al. Pult and Reide-Pult have also assessed the subjective and objective grading of meibography images. Pult and co-workers have also demonstrated that precise calculation of the area of meibomian gland loss using a computer software can lead to greatly improved repeatability in the application of grading scales. In addition, other meibomian gland software such as the width and tortuosity of the glands can be accurately computed and analyzed. Similar progress have also been reported by Srinivasan et al. These recent works demonstrated the advantages of a computerized approach to analyzing meibomian glands in meibography images.

In these recent works on computerized classification of meibomian gland loss, the images were analyzed using the image editing software ImageJ (National Institute of Health; http://imagej.nih.gov/). The user needs to be involved in identifying the gland region on the image. In other words, the user needs to tell the software where the glands are. If a large number of images is involved, this process is tedious and time consuming. Also, different examiners may draw the gland region differently, leading to inter-observer variability. Hence, it is desirable to have a method of processing meibography images that is fast, less laborious for the clinician, and is less dependent on the subjectiveness particular examiner.

A fully automated, computational approach of analyzing meibography images can address many of the difficulties mentioned, and has the potential to aid ophthalmologists in the study of meibomian glands using meibography. The purpose of this paper is to take a first step in this direction by presenting algorithms that can detect meibomian glands and classify meibography images with minimal user input. Image editing software such as ImageJ and Photoshop that require input from the user are not used. Instead, algorithms that cater specifically to the detection of meibomian glands are specially developed.

Meibography images exhibit a wide range of gland morphologies. To be of practical help to ophthalmology, one should try to sample from a wide range of images when developing the detection and classification algorithm. In particular, of most concern to clinicians is identification of images with winding glands or other complex gland patterns, because these represent cases which are intermediate between the healthy and unhealthy, and hence require the most clinical attention. However, although identification of complex gland pattern is easy for trained experts, it is a challenging task for an algorithm without

Address all correspondence to: Yang Wei Koh, Bioinformatics Institute, 30 Biopolis Street, 07-01, Matrix, Singapore 138671.
Tel.: (65) 6478 8391;
E-mail: patrickkk@bii.a-star.edu.sg

0991-3286/2012/$25.00 © 2012 SPIE
help from a human user. Hence, this paper focuses on the easiest healthy and unhealthy (criteria defined in Sec. 2.1 below) cases, two examples of which are shown in Fig. 1(d) and 1(e). There are many images of clinical interest that are neither healthy nor unhealthy according to our criteria, and we exclude these from our computational analysis in this paper.

As shown in Fig. 1(d) and 1(e), glands form a zebra-strip pattern in a healthy eye, whereas this pattern is absent in an unhealthy eye. The approach is to detect the lines along the center of and between the glands (which will be called the gland and inter-gland lines), and also the width of the glands, and then define features based on them to train a classifier to differentiate the images.

2 Methods

2.1 Subjects, Equipment, and Grading

Fifty-five patients were recruited from the dry-eye clinic and the general eye clinic of Singapore National Eye Center, including both symptomatic and asymptomatic dry-eye patients. Patients were aged 21 to 70 years. Written informed consents were obtained. The study was approved by the Singhealth Centralized Institutional Review Board and adhered to the tenets of the Declaration of Helsinki.

The patient’s chin was positioned on the chin rest with the forehead against head rest of a slit-lamp biomicroscope Topcon SL-7 (Topcon Corporation, Tokyo, Japan). This was equipped with an infrared transmitting filter and an infrared video camera (XC-ST5CE0, Sony, Tokyo, Japan). The upper eyelid of the patient was everted to expose the inner conjunctiva and the embedded meibomian glands. This is a standard procedure and causes no pain to the patient. Images were acquired using a 10x slit-lamp magnification. Care was taken to obtain the best focus as well as to avoid reflections on the inner eyelid surface. The lower eyelid was not imaged in this report because in the authors’ experience, it was easier to uniformly focus the image of the tarsal plate in the upper eyelid.

The images are manually graded by experts into 26 ‘healthy’ and 29 ‘unhealthy’ images. The clinical graders was masked to the computational classification of the images. Healthy images are those whose glands satisfy the following three criteria: 1. exhibit a zebra-like pattern, 2. are evenly long and thick, and 3. are evenly spaced and distributed along the entire eyelid margin. Unhealthy images are those that have at least 50% loss in glands in the area of interest. Images that can neither be

---

**Fig. 1** Extracting gland and inter-gland lines. (a) Lines along the centers of the glands can be located by computing the local maxima. The maxima are separated into the shown colored groups using a clustering algorithm. (b) Pixel intensity profile (after Gaussian smoothing) for one row of the image shown in (a). Gland centers are located at the local maxima. (c) Steps to process a cluster of local maxima points into a continuous curve. (d) and (e) The gland (red) and inter-glands (green) lines obtained for a healthy and an unhealthy image.
classified as healthy nor unhealthy according to the above criteria are excluded from the analysis.

2.2 Detection of Gland Length

Meibography images present difficult challenges for image processing. Apart from low contrast, nonuniformly illuminated, and out-of-focus images, frequently there are artifacts such as specular reflections and intruding eyelashes that can interfere with the detection of the glands. In this paper, a pre-processing step was taken to first manually edit away the artifacts. Then, the following computational methods are applied to detect the gland lines and widths.

2.2.1 Normalization of nonuniform illumination

Depending on the direction and focus of the infrared light source, meibography images may be nonuniformly illuminated. Figure 2(a) and 2(b) shows examples where these artifacts can be observed. If the raw image Fig. 2(a) is just enhanced using histogram equalization, the nonuniform illumination inherent in the raw image results in an image Fig. 2(b) that is bright at the center but dark at the edges. Although the stationary points (cf. Sec. 2.2.2 below) can be located in the bright region, this is difficult in the dark regions. Hence, a preprocessing step to normalize the nonuniform image intensity is needed so as to extract the stationary points in the darker regions as well. The result of normalization followed by enhancement is shown in Fig. 2(c). The stationary points obtained from Fig. 2(c) are added to those obtained from Fig. 2(b) (cf. the full algorithm in Fig. 3).

To normalize the image intensity, the nonuniform light illumination is modeled as

\[ I(x) = b(x)J(x) + n(x), \]

where \( I(x) \) is the observed image intensity at pixel \( x \), \( J(x) \) is the original signal to be restored, \( b(x) \) is the unknown nonuniform illumination field, and \( n(x) \) is the noise. The goal is to recover the original signal \( J(x) \) by estimating \( b(x) \). \( b(x) \) is estimated using fuzzy \( c \)-means clustering method and the original signal is estimated as \( J(x) = I(x)/b(x) \). After getting \( J(x) \), a 2D contrast enhancement method is applied to improve the contrast. The result is shown in Fig. 2(c).

2.2.2 Extracting gland and inter-gland lines

To detect the gland and inter-gland lines, the enhanced images [e.g., Fig. 2(c) and 2(d)] are smoothed using a Gaussian kernel (45 × 45 window, \( \sigma_x = \sigma_y = 7.25 \)). Consider the intensity profile of the smoothed image in the horizontal direction [Fig. 1(b)]. The gland centers and inter-gland points are located at the local maxima and minima of this profile, where the gradient of the pixel intensity vanishes. The colored pixels of Fig. 1(a) show that the maxima points lie along the centers of the glands. To separate the maxima points into different groups [as shown in Fig. 1(a)], where pixels belonging to the same group are represented using the same color, two maximum points are considered as belonging to the same group if they are separated by a Euclidean distance of lesser than 10 pixels. After grouping the points, to transform each group of points into one continuous line, the morphological processing steps in Fig. 1(c) are applied. The idea is to first dilate all the pixels so as to merge them into one connected component [Fig. 1(c) ii, merging], fill up any inner holes [Fig. 1(c) iii, filling], thin it to one pixel thick [Fig. 1(c) iv, thinning], and finally remove side branches [Fig. 1(c) v, pruning]. The full algorithm is summarized in Fig. 4.

Figure 4(a) and 4(c) shows the gland (maxima, red) and inter-gland (minima, green) lines obtained for a sample healthy and unhealthy image. By visual inspection, it is observed that there is a tendency for the lines in the healthy image to be longer than those in the unhealthy one. The arclength of a gland line is used as an approximation for the length of the corresponding gland, and the average arclength of all the lines (both gland and inter-gland lines) in an image is used as a feature of that image for classification.

2.3 Detection of Gland Width Using SIFT-Shannon Entropy

Scale invariant feature transform (SIFT) is an algorithm to detect and describe local features in images. For meibography images, the scale feature in SIFT is used to represent the thickness or width of the glands. Figure 4(a) shows an example of the so-called SIFT key points (red circles) computed by applying SIFT on a histogram equalized image (no normalization). The size of the key point (its scale) detects the width of the gland and the inter-gland distance [blue box in Fig. 4(a)].

Figure 4(a) and 4(b) compares the key points of a healthy and an unhealthy image. For an unhealthy image, the scales are generally smaller than those of a healthy image. The average width of all the glands in an image is approximated by the average of the scales of the key points. Let \( s^* \) be the scale of the key point \( \sigma \) on the image. Then the average scale \( \bar{s} \) is defined as

\[ \bar{s} = \frac{1}{N} \sum_{\sigma} s^\sigma, \]

where \( N \) is the total number of SIFT key points and \( \sum_{\sigma} \) means sum over all the key points.

It was also observed that for the healthy glands, because of the zebra-strip gland patterns, neighboring key points have the same scale, and are located in an orderly manner along the center of and between the glands [blue box in Fig. 4(a)]. For an unhealthy image, because there are no gland patterns, the key points are randomly scattered and have nonuniform scales. This difference in local scale distribution can be captured using Shannon entropy, which is a measure of the uniformity

\[ H(X) = -\sum_{i} p(x_i) \log_2 p(x_i), \]

where \( p(x_i) \) is the probability of the occurrence of \( x_i \). In this case, the probability distribution of the scales of the key points is used to calculate the Shannon entropy. A healthy image will have a high entropy value, while an unhealthy image will have a low entropy value. This difference in local scale distribution can be used to distinguish between healthy and unhealthy images.

\[ H(X) = -\sum_{i} p(x_i) \log_2 p(x_i), \]

where \( p(x_i) \) is the probability of the occurrence of \( x_i \). In this case, the probability distribution of the scales of the key points is used to calculate the Shannon entropy. A healthy image will have a high entropy value, while an unhealthy image will have a low entropy value. This difference in local scale distribution can be used to distinguish between healthy and unhealthy images.

By visual inspection, it was found that a threshold distance of 10 pixels gives the best classification result. Most of the groups will merge together if more than 10 pixels are used, whereas there will be many small groups if lesser than 10 pixels are used.

The observation that all the gland lines are longer in healthy images than in unhealthy ones does not hold strictly for all the images studied. To check the computational calculations, the authors drew manually and analyzed the gland lines of all the images used in this study. It was found that although the shortest glands are significantly shorter in unhealthy compared to healthy eyes, the longest ones are not significantly longer. Hence, the average length of the gland lines is used as a feature. The usefulness of this measure can partly be justified by its effectiveness in classifying the images (cf. Sec. 3).

Downloaded From: https://www.spiedigitallibrary.org/journals/Journal-of-Biomedical-Optics on 11/7/2018
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use
of a distribution. Let \( s_i \) be the scale of the \( i \)th nearest key point to the key point \( \sigma \). Define the normalized scale \( p_{\sigma i} \) for the \( i \)th nearest key point as

\[
p_{\sigma i} = \frac{(s_i)^2}{\sum_{i=1}^{n}(s_i)^2},
\]

where \( n \) is the number of nearest neighbor SIFT points considered (we use \( n = 20 \)). The Shannon entropy of the key point \( \sigma \), \( S(\sigma) \), is defined as,

\[
S(\sigma) = -\sum_{i=1}^{n} p_{\sigma i} \ln p_{\sigma i}.
\]

\( S(\sigma) \) is maximized only when all the \( p_{\sigma i} \) are equal (i.e., uniform). The average entropy of the image \( \tilde{S} \) is

\[
\tilde{S} = \frac{1}{N} \sum_{\sigma} S(\sigma).
\]

\( \tilde{S} \) should be high for a healthy image because of its uniform local distribution of scales, and low for an unhealthy one because its scales are randomly and nonuniformly distributed.

### 3 Results

#### 3.1 Separation Healthy and Unhealthy Images

Figure 5 shows how the healthy and unhealthy images are distributed according to their average arclength. The \( x \)-axis is the average arclength of the lines in an image and, to aid visualization, the \( y \)-axis shows the coefficient of variation of the lines [i.e., (standard deviation)/(mean)]. Each image is represented as a point. The healthy images (blue points) are well-separated from the unhealthy images (red points) along the \( x \) direction, meaning that the average arclength is a good feature for separating the healthy and unhealthy class. Figure 5 shows the distribution of the healthy and unhealthy images according to their \( \tilde{s} \) (\( y \)-axis, average scale) and \( \tilde{S} \) (\( x \)-axis, average entropy). Once again, the healthy and unhealthy points are well-separated, meaning that they are also good features.

#### 3.2 Accuracy of Classification

The three features (average arclength, \( \tilde{s} \), and \( \tilde{S} \)) will be used in the next section as features for classification.
healthy and 15 unhealthy) are randomly selected as training data set to train the SVM, and the remaining half (13 healthy and 14 unhealthy) are used as the testing set. The success rate, defined as the number of times the SVM correctly predicts the correct label (i.e., healthy or unhealthy), is calculated for both the training set and testing set. This is repeated 100 times, where each time a different set of training and testing images are chosen randomly. The success rates are averaged over the 100 repetitions, and the results are shown in Table 1. For the training data, the classifier has an average specificity of 97% for predicting the healthy images, and sensitivity of 100% for predicting the unhealthy ones. For the testing data, the result is a specificity of 96% and sensitivity of 98%.

Table 1 Success rate of predicting the correct label (healthy or unhealthy) using a linear SVM. The results are obtained by averaging over 100 different random selections of training and testing data sets.

<table>
<thead>
<tr>
<th></th>
<th>Specificity (± stand.err.) (%)</th>
<th>Sensitivity (± stand.err.) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training data</td>
<td>96.8 ± 0.4</td>
<td>100.0 ± 0.0</td>
</tr>
<tr>
<td>Testing data</td>
<td>96.1 ± 0.4</td>
<td>97.9 ± 0.6</td>
</tr>
</tbody>
</table>

4 Discussion
In this paper, meibography images are classified using criteria based on gland lines and width, which are different from the well-established criteria based on the area of meibomian gland loss. In order to compute the area of meibomian gland loss, the gland region needs to be segmented from the background nongland region. Without any input from the user, this is computationally challenging because the image pixel intensity changes gradually between the borders and edges of the glands, making it difficult for the algorithm to decide the appropriate boundaries between glands. Gland lines and width, on the other hand, are easier to detect without the user, and the results presented in this paper can be considered as successful first steps based on these simpler features. In future work, these two features will be combined to segment the individual glands. This will allow the gland area to be computed, which will then allow images to be graded according to the established criteria.

This study focused on images of the upper eyelids because it was easier to obtain a uniformly focused image of the tarsal plate. Clinically, however, it is sometimes easier to work with the lower lids because they are less uncomfortable for the patients. The methods presented here have also been tested on 10 healthy lower lid images. It was found that the gland and inter-gland detection algorithm is effective for the lower lid images. For the detection of gland widths, the bright and
dark strips of the zebra-strip pattern are different from the upper lids, with the dark strips being very narrow and the bright strips being much wider and frequently touching. Width detection for the bright and dark strips should be done separately, and using different scales. The calculated features for the lower lids are also different compared to the upper lids, so classification of upper and lower lids should be performed separately. Overall, the techniques here are applicable to both upper and lower lids.

In addition to gland length and width, meibomian glands can also be characterized by other morphological features such as shape, contour, and tortuosity. Such features are important when grading images that are intermediate between the healthy and unhealthy ones. A limitation of the current study is that these features were not assessed. The healthy and unhealthy glands images considered here are extreme cases, and, although easy for any trained examiner, is challenging for an automated algorithm. This paper focuses on them because any algorithm working without user input should first be able to classify these simplest cases before addressing more difficult intermediate cases. It is shown here that based on the detected gland lines and width, the classification performed by the algorithm agrees closely with that done by human experts.

When applied to the classification of intermediate images, the current algorithm utilizing only gland length and width failed to produce satisfactory result. Instead of being distinctly separated in feature space like the healthy and unhealthy cases (Fig. 3), the intermediate cases overlap with the healthy and unhealthy points. This makes it difficult for the SVM to classify. However, this is expected because studies have already shown that the meibomian gland structure of intermediate cases must be described by additional morphological features such as tortuosity, whereas in the current computational method, only the length and width of the gland are considered. To treat the intermediate cases, additional morphological features can be computed based on the gland lines already obtained, and be used to expand the dimension of the feature for the SVM. The assessment of these morphological features and their application to the grading of intermediate images will be reported in future work.

An important difficulty encountered in this work concerns how images are taken. Slightly different ways of taking an image will not influence the gland structure. But because it introduces complication during the computational analysis, the final conclusions might be affected. This problem is important in this paper because the objective is to have a user-free assessment of the gland structure, so any error caused by detection of erroneous artifacts present in the image will not have the chance to be corrected by the user. This ultimately introduces errors into the final results. To be specific, there may be artifacts such as intruding eyelashes, specular reflections from the tear film, and misalignment of the gland region in the image. In this paper, the primary concern is to demonstrate the ability of the proposed algorithms in detecting meibomian glands, and so the computational difficulty of handling artifacts was dealt with heuristically. An image editing software was used to edit away parts of the image not directly related to the gland region by manually drawing out the region of interest in an image. The gland detection algorithm are then applied to the identified region. Although this process of drawing out gland region manually may appear similar to what has already been done in previous studies, the important difference is that here the region inside the identified area is subjected to gland detection by the algorithm, whereas in previous studies the area of interest are drawn by the user with the intention of locating the glands.

As the ultimate objective is to achieve completely automated detection of gland regions without any user input, it is important to also develop an algorithm for locating the area of interest. It is found that this can be achieved by a simple change in the imaging protocol. The magnification when taking an image should be lowered such that the upper eyelid margin and the edge of the upper tarsal plate are both visible on the image; the margin and edge can then be used as references lines to define the gland region. This two lines can be detected using image processing techniques and then the area of interest is obtained. Such a user-free way of locating the area of interest is desirable because it eliminates the subjectiveness introduced by clinicians when drawing the area of interest. Different examiners may draw the area differently; an algorithmic method, on the other hand, will always produce the same region everyday. In the current study, images analyzed were collected from an earlier clinical study where the need to include the upper eyelid margin and upper tarsal plate in the images was not foreseen. In subsequent studies, this precaution will be taken and the area of interest will be automatically identified. This work will be reported in a future paper.

5 Conclusion

In this paper, a use-free computational approach to detecting meibomian glands and classifying meibography images was presented. The gland and inter-gland lines, and gland width were detected algorithmically, and then used as features for classification of the images. The classification results by the computational approach agrees closely with that done by human experts, with a specificity of 96% for healthy images, and sensitivity of 98% for unhealthy ones.

Acknowledgments

We thank Ivy Law and Choon Kong Yap for their comments. This work is supported in part by the Agency for Science, Technology, and Research (A*STAR) of Singapore, Biomedical Research Council/Translational Clinical Research Programme 2010 Grant 10/1/06/19/670, and National Medical Research Council individual grants NMRC/1206/2009, NMRC/CSA/013/2009, and NMRC/CG/SERI/2010.

References